Abstract

Workers’ compensation cases are often contentious, litigated events. The authors describe a Gordian Knot that is created by competing interests. These interests are manifested in secondary gain issues, non-physical factors (psychological, social, financial and economic) and the expert witness culture. The process of making medical, legal and case management decisions is compromised by the failure to recognize the limitations of the four primary components of the most commonly administered Functional Capacity Evaluation (FCE) protocols. Three of these components (pain questionnaires, intake interview and clinical assessment) are discussed in this first part of
a two-part article. This discussion lays the foundation for Part 2, which addresses the fourth FCE component - the most commonly administered tests which are used for the alleged purpose of classifying validity of effort in Part 2. The authors provide insights into how the Gordian Knot can be cut with more accurate, statistically-based testing methods described in peer reviewed publications.

Introduction

One of the central issues in any compensable injury case is related to the functional capacities of the claimant. Can the claimant return to full duty? Can the claimant return to modified duty? Can the claimant take part in any occupational activity of any kind? These are questions that must be answered to make proper medical, case management and legal decisions. It is tempting to say that the act of answering these questions can come about only by untangling a Gordian Knot of competing interests and a wide array of factors, some of which are physical and some of which are non-physical - all of which can affect the applicant as well as the defense. Human nature has not changed since Alexander the Great’s legendary “solution” to solving a complex problem. We seem to be preoccupied with the challenge of untangling a knot, when a direct approach to solving a complex problem might prove more effective. Alexander’s direct approach, in fact, is an object lesson for persons who are attempting to untie a modern Gordian Knot, known as “workers’ compensation.” In this, the first of a two-part article, we will present the case for cutting the knot by answering just one basic question: Did the claimant cooperate during a functional capacity evaluation (FCE)? We will show how standard testing methods have failed to answer this question. Lastly, we will bring forward information regarding new and accurate testing methods, which will cut through the many factors which have historically complicated the case management process.

FCEs have been a staple service in the field of industrial rehabilitation since the field’s inception more than 30 years ago. The purpose of an FCE is to provide information relevant to answering the ques-
tions introducing this paper. FCEs are commonly ordered to aid in the evaluation of an insurance claim of full or partial disability. An FCE typically involves a number of measures of physical work capacity. A major aspect of such an evaluation concerns validity of effort. Put simply, is the claimant performing up to his or her true capacity in the test? In an unknown, but probably fairly large proportion of FCEs, the claimant is faking (or exaggerating) an injury or disability. It should be obvious that the results of an FCE are only useful if the claimant gives a valid effort. Determination of validity of effort is therefore of paramount importance. If the validity of the claimant’s performance is not objectively and definitively assessed in the FCE, the FCE is of no value because the quantification of function is utterly dependent upon an accurate assessment of validity of effort. Note that we refer to “validity” of effort throughout. The term “sincerity” of effort is also widely used and synonymous.

Unfortunately, making medical and legal case management decisions is complicated by the fact that diagnoses, even when they are correct, do not predict outcome. After all, some persons with low back strains never return to work, while others who have multiple-level fusions may go back to perform physically-demanding jobs. What factor or factors are responsible for such divergent outcomes? Most likely, that is an unanswerable question that would require quantifying the effects and importance of psychological, social, economic and financial variables that do not lend themselves to precise and objective measurement. Certainly these are factors that are difficult to take into account in making case management decisions. The ability of medical professionals and claims administrators to effectively address or even objectively measure such issues is limited.

Enter the expert witness. The most commonly used methods of assessing validity of effort are actually a series of observations and measurements that are “interpreted” by the test administrator. When experts disagree, attorneys argue the relative merits of interpretations in court. In both civil and workers’ compensation systems, there is an expert witness culture in which one witness’ testimony (an “interpretation”) is contradicted by an expert witness that has
been hired by the opposition. A trier of facts, administrative panel or jury makes a judgment. Such judgments are arguably affected by the relative skills of the attorneys, the impressions made by the expert witnesses, and the ability of those making the decision to “interpret” the testimony, apply the law and render an impartial decision. Thus, in disputed cases, questions relative to the functional abilities and validity of effort are settled by contests between dueling experts, not objective findings of an FCE. This is not acceptable. Even a researcher who has published multiple studies on the “visual estimation of effort” (methodology discussed in length in future Part 2 of this article) wrote of the importance of the “evidence” versus the “expert witness” (Reneman, 2003).

The issue of “secondary gain” compounds the contentious nature of compensable injury cases, and relates to the fact that diagnosis does not always predict outcome. It is not uncommon to hear about the “injured” worker who is caught on video working on another job, running a marathon or otherwise being caught in the act of committing fraud. As a result of these relatively dramatic instances workers with legitimate claims are often tarred with the same brush as the fraudulent claimant.

It is easy to forget that possible “secondary gain” may also affect the actions of other parties involved in compensation cases. By this we mean that a party’s conduct or representations may be unduly influenced by financial self-interest. For instance, some care providers may be tempted to advocate for (and gain economically from) treatment that is unnecessary, and do so with no objective foundation for believing the patient has a functional deficit. Others might foresee potential economic gain by making statements or taking actions that will enhance the loyalty of a guarantor or employer by obstructing justifiable treatment for an injured worker. Furthermore, it should be pointed out that employers and guarantors, who have an obvious financial stake in the outcome of a case, could conceivably be affected by secondary gain considerations. A robust assessment of validity of effort, though, would not only assist in quantifying function, but would also shine a very bright light, not only on the claimant, but
also on all of the players in the medical field as well as the employers and guarantors.

**FCE In The Claims Process**

The FCE represents a crucial moment in the claims management process. Obviously, an FCE is performed as an “end game” tactic, primarily employed when cases become protracted. The FCE is intended to establish a consensus for future case management decisions. Arguably, the FCE is an event during which all of the effects of various psychological, social, economic and financial variables on both sides of the claim intersect - or rather collide. It is at this time that the test behavior of the claimant may enable a test administrator to provide valuable insights into the claimant’s test behavior. Test behavior is arguably impacted, negatively or positively, by the interaction between the non-physical factors and medical and functional variables. If test behavior indicates a valid effort, then the test has also most likely quantified functional capacities. If test behavior indicates an invalid effort, the treatment approach can be reassessed and modified to take these psycho-social factors into account, or the case can be litigated. Neither normal nor abnormal test behavior, unfortunately, can be quantified using the “standard” validity of effort testing methods. Again, if validity of effort cannot be established, functional capacities remain unknown. Thus, any recommendation with regard to case management is based on unproven assumptions.

Given the potential utility of an FCE, it is disconcerting at the least to know that professionals in physical therapy have been given ample warning of the failure to objectively address the issue of validity of effort. Editorials in two widely read trade publications have brought the issue to the forefront: the issues of validity of effort and the importance of using testing methods that are “evidence-based” (Simons, 2006; Mecham, 2008). The former editorial was written from the perspective of a physical therapist - who also happened to be an attorney. The title of her opinion piece, “Credibility Crisis in FCEs,” is itself instructive. With the question of credibility of the testing
process looming large, one would assume that practitioners would be concerned with the admissibility of “evidence” that is presented in medical-legal cases, particularly in light of the landmark Supreme Court decision *Daubert v Merrill Dow*, 509 U.S. 579 (1993).

The Daubert decision established standards for the admissibility of evidence in medical-legal cases. The Daubert decision originally applied to federal cases. However, it has also been applied to workers’ compensation cases such as *Estate of George v. Vermont League of Cities and Towns*, 2010 VT 1 (Jan. 15, 2010) and *Vannoy-Joseph v. State/DMHAS*, 5164 CRB-8-06-11 (January 29, 2008). At least 17 states have adopted the Daubert standard outright and 15 more have adopted standards similar to Daubert.

In *Daubert*, the court defined the term “scientific knowledge” and said that evidence introduced into medical-legal cases could be ruled inadmissible if it does not meet the standard set by the court. It is not the purpose of this presentation to argue the merits of *Daubert* or to give the reader legal advice pertaining to the decision or to the degree that it might be applied to the subject matter of this paper. We are simply pointing out an important legal precedent and the fact that its implementation in various legal jurisdictions has been gradually widened since 1993. The standard has not, to the knowledge of the authors, been applied to FCEs - yet. But a *Daubert* ruling relative to FCE assessments of validity of effort in some state seems inevitable. In most cases, there is either little or no published evidence that standard validity of effort testing methods is effective. Worse, there is credible evidence that documents their inaccuracy, discussed below.

While there is usually a wealth of information in the usual FCE report, we question whether much of the information is useful, either in the classification of function or validity of effort. We believe that there are few opportunities to objectively assess validity of effort during an FCE. That is not to say that there is a lack of data, which are represented as or presumed to be “objective” indices of function and validity of effort. Uncritical acceptance of many common measure-
ments and observations will inevitably result in the inclusion of inaccurate assessments of function and validity of effort into the medical and legal record. Therefore, some basic information pertaining to such measures and observations is in order.

The FCE process usually consists of four components:

1. Pain questionnaires.
2. Intake interview.
4. Functional testing.

Pain questionnaires reveal nothing about function or validity of effort, although “high” scores on various instruments may be predictive of compliance or non-compliance during a test, or failure to improve with treatment. But the predictive value is only the chance or odds of an individual with a range of pain score behaving in a certain way or having specific psychological traits that can be identified with extensive psychological testing. Examples of pain questionnaires include: Oswestry Low Back Inventory, Modified Somatic Perceptions Questionnaire and the 0 – 10+ Pain Scale. These instruments do not, however, predict cooperation or non-cooperation for individuals undergoing functional assessment (Schapmire, St. James, Feeler & Kleinkort, 2010). Written instruments may provide insights into various behavioral or psychological features of a given individual. Many times, they are administered with the intention of identifying “symptom magnification.” However, there is no universal agreement on the definition of the term. In fact, there is no universal agreement that the term is even useful. We make a clear distinction here between short pain questionnaires and more lengthy and formal psychological tests such as the Minnesota Multi-Phasic Personality Inventory. It is noted that even with much more sophisticated psychological testing, there is still intense debate over the applicability of such instruments in the compensation population. However, a discussion of psychological tests such as the MMPI falls beyond the purview of this article.
During an intake interview it is sometimes possible to discover new information that will perhaps be useful during the testing process or in making recommendations. The obvious limitation of an intake interview is that the quality and reliability of the responses to an interviewer’s questions is unknown. Furthermore, the quality of information obtained during an interview is also affected by the skills of the person conducting the interview. While the subjective reports may in fact be accurate, they are subject to challenge unless they are substantiated at some point in time by objective data. It is a mistake to consider self-reported dysfunction as “objective” on face value. Generally, information obtained during an intake interview, from pain questionnaires, or from a clinical assessment should be regarded with skepticism unless it is correlated with multiple objective data points. Widespread lay access to medical information, via the Internet, is a potential confounding variable, the importance of which should not be discounted out of hand.

Clinical assessments involve the administration of various clinical tests and an interpretation of the test results. There are two primary problems with clinical tests for orthopedic dysfunction. First, there is a problem related to standardization in administering the tests. To illustrate this point, we suggest that you go to www.Youtube.com. Then conduct a search, using the name of nearly any orthopedic test as the search term (Thomas Test, Straight Leg Raise, Adson’s Test, Anterior/Posterior Drawer Sign, etc.). You will discover that there are multiple methods that are purported to be the “correct” way to administer just about any orthopedic test. Second, most orthopedic tests rely on an accurate answer from the patient for proper interpretation. Nearly all clinical tests are interpreted on the basis of whether the claimant indicates the maneuver either increases or decreases symptoms. Therefore, accuracy in the claimant’s response is critical to drawing proper conclusions. Assuming that any subjective report is accurate may lead to false conclusions related to the clinical tests that have been administered. Assuming that all subjective reports by the claimant are inaccurate may also lead to incorrect conclusions regarding physical status.
Many times, a series of clinical tests are conducted. In these cases, it is assumed or believed that the tests should somehow correlate with one another - sometimes in a vague or even questionable way. But if all of the responses “correlate” with one another, then it is assumed that accurate conclusions can be made regarding physical status. Even this approach is subject to challenge, given the absence of standardization in test administration and the limitation relative to the accuracy of the claimant’s response. Unless the clinical tests have been somehow cross-validated with objective testing, the results of any clinical assessment can be successfully challenged even if the conclusions arising from the clinical tests are correct.

Observations and Physical Measurements

During the FCE, there are two types of data, which may or may not be meaningful in the assessment of function and the classification of validity of effort. First, there are observations. Second, there are physical measurements.

Some observations can be extremely objective and very meaningful. Inconsistencies in performance allow for the logical conclusion that validity of effort is questionable. For example, if a gait deviation migrates from one side to the other, accurate conclusions can be drawn from the presentation. Likewise, if ranges of motion for, say, a rotator cuff patient, are significantly different during various parts of the FCE, including the clinical assessment, objective and accurate conclusions can likewise be made. Inconsistencies of these kinds strongly suggest an invalid effort.

Obviously, inconsistencies between repeated measures logically demonstrate the physical performance data are not valid. But a common mistake in interpreting clinical tests is to assume that consistency automatically means that the data are, in fact, valid. This is an obvious error in logic. Consistency in performance does not automatically confer the mantle of “valid.” If a “consistent” performance is observed during repeated observations with distraction-based
tests, *validity might be assumed*, provided the preponderance of *objective* measures of validity. The opportunity and feasibility of developing valid distraction-based tests for all observations of this kind, though, is limited.

Observations which may or may not be objective, accurate or relevant in the assessment of function and validity of effort include statements pertaining to facial affect, muscular tension, “cogwheeling” during manual strength testing, and the set of five different observations involved the administration of the well-known Waddell Tests. The accuracy of such observations is completely dependent upon the skill and experience of the evaluator. Hence, these are not standardized methods of assessment and can be effectively challenged, even when they are “correct.” Observations related to physical changes that are felt, heard or visually observed during a clinical assessment or during functional testing fall into the category of “semi-quantitative.” Included in this group are:

1. Observations relative to color and/or color changes, e.g., “bluish” or “increased redness.”
2. Observations relative to surface temperatures, e.g., “hot” v. “cold.”
3. Observations relative to muscle tone, spasm or tension.
4. Observations relative to crepitus during movement of various joints.
5. Observations relative to posture, e.g., “forward head posture,” “rotation to the right,” and “normal lordosis.”
6. Observations relative to ilial crest height or leg length differences (unless they are physically measured).
7. Observations regarding the relative laxity in a joint, e.g., perceived anterior or posterior translation of the knee during the Anterior/Posterior Drawer Test.

The utility of these observations with regard to validity of effort is questionable, given the fact that they do not involve taking physical
measurements (the possible exception of No. 6, above). Their relation to function is also challengeable.

Some common components of FCEs are observations that may be “substantiated” by documenting the time of participation in an activity, repetitions performed, distance traveled, or weight that is lifted. Hence, they are combinations of observations and physical measurements. Examples include:

1. “Job simulation” testing. Such activities may include information related to the amount of time for the testing, the amount of weight lifted during the task and the number of repetitions performed.

2. Positional tolerance testing. A typical assessment of this kind might evaluate the ability to remain in a deep squat posture for a given period of time. Another assessment might be activity in which the claimant was able to work with the arms above the head.

3. Tolerance for repetitive movements. Examples of such a test might be a tally of the number of times the claimant was “able” to repetitively bend at the waist, or the number of times an individual is able to repetitively flex the shoulder.

The activities on this list above face the same challenge as any other observations. They are not valid indices of effort - simply because the outcome is dependent upon the cooperation of the subject. Furthermore, there are no established standards for distinguishing between “normal” and “abnormal” function for such tests. There is no supportive literature relative to classifying validity of effort during these activities. The most that can possibly gained from these activities is documentation of functional abilities under ideal conditions for an arbitrarily chosen period of time - but only if a robust assessment of validity of effort has been passed during activities for which validity of effort can actually be measured. Even then, validity of effort during job simulation, positional tolerance testing and tolerance for repetitive movements can only be presumed.
There are a small number of tests outside of functional testing which involve actual physical measurements. These include:

1. Volumetric measurements of the extremities.
2. Circumferential measurements.
3. Temperature measurements using a thermometer.
4. Range of motion testing.

Volumetric, circumferential and temperature measurements are objective. Their importance, however, may be of questionable value unless they show striking deviations from established norms or reveal significant asymmetry when comparing the right and left sides of the body. While stark differences between one side of the body and the other may very well indicate an organic problem, such differences do not, in and of themselves, speak to the issues of function or validity of effort. There is, in fact, a lack of empirical evidence that links “abnormalities” in volumetric, circumferential and temperature measurements to function abilities.

Even range of motion testing, the most basic measurement made in the field of physical therapy, is fraught with challenges. One such challenge is that when using a mechanical goniometer (protractor), placement of the hinge on the testing device must be properly aligned with the axis of rotation of the joint - and both sides of the device must be accurately aligned with the parts of the body that are being assessed. Small differences in placement with regard to the axis of rotation or the alignment of the device relative to the body will result in measurement error. Furthermore, there are no established norms for “normal” and “abnormal” ranges of motion - and there is no proof that ranges of motion are related in any meaningful way to function.

Although some editions of the American Medical Association’s Guides to the Evaluation of Permanent Impairment use ranges of motion in calculating “impairment,” even these norms are unsupported by any published study. In addition, although repeated measurements are
used for the alleged purpose of assessing cooperation during range of motion testing, there is absolutely no published study that defines “acceptable” and “unacceptable” variation between trials. Lastly, even to the casual lay observer, it should be obvious that purposely restricting ranges of motion for the back, for example, is fairly easy for an individual who is determined to misrepresent true range of motion.

All that a test administrator can hope to provide when making the aforementioned observations and measurements is a very general impression with regard to cooperation - and such impressions are extremely vulnerable to attack, particularly when they are not correlated with physical performance data that can actually be analyzed for validity of effort. If the testing process is comprised solely of measurements and observations such as those discussed above, legal challenge to the methodology is a straightforward process for an experienced attorney.

Physical performance testing during functional testing represents the only part of an FCE in which actual physical measurements are taken and subsequently analyzed for validity of effort. Even then, the activities that lend themselves to an accurate and objective analysis of effort are limited. Typically, these activities fall into three categories:

1. Hand strength assessment.
2. Strength assessments using isokinetic or isometric equipment.

Unfortunately, “standard” testing methodology for validity of effort testing during hand strength assessments are inappropriate for classifying validity of effort according to the conclusions of many published studies and reviews (Ashford, Nagelburg, & Adkins, 1996; De Smet & Londers, 2003; Dvir, 1999; Fairfax, Balnave, & Adams, 1995; Fishbain, Cutler, Rosomoff, & Rosomoff, 1999; Goldman, Cahalan, & An, 1991; Goldman et al., 1991; Guitierrez & Shechtman,
Various types of “machine testing” have been promoted as - but never been shown to be - highly effective in predicting dynamic function or assessing validity of effort. In fact, major reviews have called isokinetic testing into question (Rothstein, Lamb, & Mayhew, 1987; Dvir, 1991; Newton & Waddell, 1993). Furthermore, new research demonstrates conclusively that isometric testing is not appropriate for use in quantifying function or validity of effort (Feeler, St. James, & Schapmire, 2010; Townsend, Schapmire, & St. James, 2010).

The visual estimation of effort (VEE) has been promoted for many years as a way to assess a claimant during a lifting evaluation with an array of correlation studies. These studies will be cited and discussed in Part 2 of this article.1 Recent research has demonstrated that the VEE approach is highly inaccurate and likely does not meet the Daubert standard for the admissibility of evidence (Schapmire, St. James, Townsend, & Feeler, In Press, scheduled for publication in 2011).

Conclusion

It is understandable the reader may question the comments relative to “standard” testing methodology in the three previous paragraphs. After all, the most widely used validity of effort tests have been accepted and use for more than 30 years. Like the Gordian Knot though, these methods are based largely on myth. Like all myths, they have been accepted because they have a plausible basis; and like

---

1 Editor’s note: to be published in the next edition of the Journal.
all myths, they have not been - and can never be - substantiated as being accurate. Part 2 of this article will specifically address the three most common myths in validity of effort testing, explain the plausible - but unjustified - bases for their acceptance and provide ample reason to reject their use. Cost-effective methods that are supported by peer review will be proposed as the best possible tool to cut the modern Gordian Knot we know as “workers’ compensation.”

References


Darrell Schapmire, MS (exercise physiology), is a 1991 graduate of Benedictine University, Lisle, IL. He is a partner in X-RTS Software Products, Inc. and sole proprietor of Industrial Rehabilitation Consultants. In addition to performing functional capacity evaluations, he has patented two proprietary devices and worked closely with his business partner, James St. James, to conduct research focusing on validity of effort testing.

James D. St. James, PhD (experimental psychology), is a full-time faculty member of Millikin University, Decatur, IL. In addition, he is a partner in X-RTS Software Products, Inc. He and Mr. Schapmire co-developed the concept of simultaneous bilateral testing of the hands for validity of effort testing. They have worked closely on several successful research projects that have brought new insights to the field for the purpose of objectively classifying test behavior of insurance claimants.